PROCEEDINGS OF TDWG

Biodiversity Information Science and Standarts 1: e20232
doi: 10.3897/tdwgproceedings.1.20232

OPEN ACCESS

Conference Abstract

TAXREF-LD: A Reference Thesaurus for Biodiversity on the Web of Linked Data

Franck Michel‡, Catherine Faron-Zucker‡, Sandrine Tercerie§, Gargominy Olivier§

‡ Université Côte d'Azur, Inria, CNRS, I3S, Sophia Antipolis, France
§ Muséum national d'Histoire naturelle, Paris, France

Corresponding author: Franck Michel (franck.michel@cnrs.fr)

Received: 12 Aug 2017 | Published: 14 Aug 2017

Citation: Michel F, Faron-Zucker C, Tercerie S, Olivier G (2017) TAXREF-LD: A Reference Thesaurus for Biodiversity on the Web of Linked Data. Proceedings of TDWG 1: e20232.
https://doi.org/10.3897/tdwgproceedings.1.20232

Abstract

Started in the early 2000's, the Web of Data has now become a reality [Bizer 2009]. It keeps on growing through the relentless publication and interlinking of data sets spanning various domains of knowledge. Building upon the Resource Description Framework (RDF), this new layer of the Web implements the Linked Data paradigm [Heath and Bizer 2011] to connect and share pieces of data from disparate data sets. Thereby, it enables the integration of distributed and heterogeneous data sets, spawning an unprecedented worldwide knowledge base.

Taxonomic registers are key tools to help us comprehend the diversity of nature. They are the backbone for integrating independent data sources, and help figure out strategies regarding biodiversity and natural heritage conservation. As such, they naturally stand out as potential contributors to the Web of Data. Several international initiatives on taxonomic thesaurus such as NCBI Organismal Classification [Federhen 2012], AGROVOC Multilingual agricultural thesaurus [Caracciolo et al. 2013] or Encyclopedia of Life [Blaustein 2009] have already made this move towards the Web of Data.

In this talk, we will present an on-going work related to TAXREF [Gargominy et al. 2016], the taxonomic register for fauna, flora and fungus, maintained and distributed by the National Museum of Natural History of Paris (France). TAXREF registers all species

© Michel F et al. This is an open access article distributed under the terms of the Creative Commons Attribution License (CC BY 4.0), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

inventoried in metropolitan France and overseas territories, in a controlled hierarchy of over 500.000 scientific names. Our goal is to publish TAXREF on the Web of Data, denoted TAXREF-LD, while adhering to standards and best practices for the publication of Linked Open Data (LOD) [Farias Lóscio et al. 2017].

The publication of TAXREF-LD as LOD required tackling several challenges. Far beyond a sheer automatic translation of the TAXREF database into LOD standards, the key point of the reported endeavor was the design of a model able to account for the two coexisting yet distinct realities underlying TAXREF, namely the nomenclature and the taxonomy. At the nomenclatural level, each scientific name is represented by a concept, expressed in the Simple Knowledge Organization System (SKOS) vocabulary [Miles and Bechhofer 2009], along with an authority and a taxonomic rank. At the taxonomic level, a species is represented by a class in the Web Ontology Language (OWL) [Schneider et al. 2012] whose properties are the species traits (habitat, biogeographical status, conservation status...). Both levels are connected by the links between a species and associated names (the valid name and existing synonyms). Note that the modelling applies not only to species but also to any other taxonomic rank (genus, family, etc.).

This model has several key advantages. First, it is relevant to biologists as well as computer scientists. Indeed, it agrees with three centuries of thinking on nomenclatural codes [Ride et al. 1999, McNeill et al. 2012] while, at the same time, it fits in with the philosophy underpinning SKOS and OWL: the nomenclatural level allows circulating through a hierarchy of concepts representing scientific names, and at the taxonomic level, the OWL classes represent the sets of individuals sharing common traits. Second, the model enables drawing links with other data sources published on the Web of Data, that may represent either nomenclatural or taxonomic information. Third, the taxonomy evolves frequently along with newly discovered species and changes in the scientific consensus. Typically, a name may alternatively be considered as the valid name of a species or a synonym. The distinction between the nomenclatural and taxonomic levels, alongside an appropriate Uniform Resource Identifier (URI) naming scheme for names and taxa, makes the model flexible enough to accommodate such changes.

Furthermore, our goal in this talk is not only to present the work achieved, but more importantly to engage in a discussion with the stakeholders of the community, may they be data consumers or producers of sibling classifications concerned with the publication of LOD, about data integration scenarios that may arise from the availability of such a large, distributed, knowledge database.

Keywords

Linked Data, Taxonomy, Data Integration

Presenting author

Franck MICHEL is a research engineer at the University Cote d'Azur, France. His research topics notably concern the integration and federation of heterogeneous data sources using Semantic Web ontologies, and their publication in the Web of Data.

Olivier GARGOMINY is a research engineer at the National Museum of Natural history in France. He is responsible for the French national taxonomic register for fauna, flora and fungus (named TAXREF) and the knowledge database associated with this taxonomic register (status, biological interactions, etc).

References

- Bizer C (2009) The Emerging Web of Linked Data. IEEE intelligent systems 24 (5): 87-92.
- Blaustein R (2009) The Encyclopedia of Life: Describing Species, Unifying Biology. BioScience 59 (7): 551-556.
- Caracciolo C, Stellato A, Morshed A, Johannsen G, Rajbhandari S, Jaques Y, Keizer J (2013) The AGROVOC linked dataset. Semantic Web 4 (3): 341-348.
- Farias Lóscio B, Burle C, Calegari N (2017) Data on the Web Best Practices. W3C Recommandation.
- Federhen S (2012) The NCBI Taxonomy database. Nucleic Acids Research 40 (D1): D136-D143-D136-D143.
- Gargominy O, Tercerie S, Régnier C, Ramage T, Schoelink C, Dupont P, Vandel E, Daszkiewicz P, Poncet L (2016) TAXREF v10. 0, référentiel taxonomique pour la France: méthodologie, mise en œuvre et diffusion. Muséum national d'Histoire naturelle, Paris. Rapport SPN 2016.
- Heath T, Bizer C (2011) Linked Data: Evolving the Web into a Global Data Space. 1st. Morgan & Claypool
- McNeill J, Barrie F, Buck W, Demoulin V, Greuter W, Hawksworth D, Herendeen P, Knapp S, Marhold K, Prado J, Prud'homme Van Reine W, Smith G, Wiersema J, Turland N (Eds) (2012) International Code of Nomenclature for algae, fungi, and plants (Melbourne Code). International Association for Plant Taxonomy.
- Miles A, Bechhofer S (2009) SKOS Simple Knowledge Organization System Namespace Document. W3C Recommendation.
- Ride WD, Cogger HG, Dupuis C, Kraus O, Minelli A, Thompson FC, Tubbs PK (Eds) (1999) International Code of Zoological Nomenclature. Fourth edition. International Trust for Zoological Nomenclature
- Schneider M, Carroll J, Herman I, Patel-Schneider P (2012) OWL 2 Web Ontology Language RDF-Based Semantics (Second Edition). W3C Recommendation.